A diszfónia súlyosságának automatikus becslése a szakértői értékelések szubjektív jellegének figyelembevételével

Tulics Miklós Gábriel[1], Jászai Henrietta[1], Vicsi Klára[1]

[1]Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
{tulics,vicsi}@tmit.bme.hu
{jaszaiheni}@gmail.com

Kivonat Ebben a munkában egy a diszfónia súlyosságát automatikusan becslő eljárás kerül bemutatásra. A k-közép algoritmus segítségével kimutattuk, hogy a különböző fonetikai osztályokba tartozó beszédhangokon mért akusztikai paraméterek alkalmasak a szakemberek négyfokozatú értékelésének (RBH szubjektív skála) modellezésére. A kapott felügyelet nélküli automatikus becslési eredmény meglepően közel áll a szubjektív megítéléshez. Ezen kívül megvizsgáltuk négy szakember értékelésének következetességét. A négy szakember egyike állította fel a diagnózist és értékelte a beteg beszédének minőségét a konzultációk során; a másik három szakember nem ismerte a pácienst, és csak a korábban rögzített hangfájlok alapján határozta meg a diszfónia súlyosságát. A diszfónia súlyosságának automatikus becslése során két regressziós modell került összehasonlításra: a négy szakember értékelésének átlaga alapján készített és a betegeket kezelő klinikus ítélete alapján létrehozott modell. A két modell alacsony RMSE és magas korrelációs értékeket ért el az automatikusan becsült súlyosságot és a perceptuális értékeléseket összevetve.
Kulcsszavak: beszédelemzés, folyamatos beszéd, kóros beszéd, következetesség vizsgálat, regresszióelemzés, klaszterelemzés, diagnosztika

1. Bevezetés

A fejlett országokban a dolgozók egynegyedénél elengedhetetlen a beszéd használata kommunikációs eszközként. A beszédhang-képzés bármilyen rendellenessége kihatással van a magánéletre, érintheti a szakmai előmenetelünket és a megélhetésünket is. Új, olcsó és hatékony módszerekre van szükség, amelyek segítik a háziorvosok, fül orr gégészek munkáját a betegségek feltárásában.

A kutatók célja egy olyan automatikus becslési rendszer létrehozása, amely felismeri a kóros hangot, és automatikusan meghatározza a diszfónia súlyosságát, ezáltal lehetőség nyílik a hang non-invazív és objektív diagnosztizálására, ezzel együtt a korai felismerésre. A diszfónia a hangképzés diszfunkciójára utal. A diszfóniás hang rendszerint rekedt, levegős, fátyolos, préselt vokális tulajdonságokkal jellemezhető [1]. A folyamatos beszéd vizsgálatának számos előnye van

a kitartott magánhangzók elemzésével szemben, mivel a folyamatos beszéd tartalmaz számos alapfrekvencia variációt, szüneteket és lehetőség van a beszédhangok különböző változatainak elemzésére. A kutatási eredmények várhatóan jobban alkalmazhatók a gyakorlatban, mivel a valós életben a folyamatos beszédet használjuk [2,3]. Az egészséges és kóros hangok automatikus osztályozásában az eredmények jelentősen javultak, amikor folyamatos beszédet kezdtek el használni a kitartott magánhangzók helyett [4].

A kutatók olyan akusztikai paraméterek kifejlesztésére összpontosítottak, amelyek hatékonyan reprezentálják a hangképző rendszer kóros állapotát. A legszélesebb körben használt akusztikai paraméterek közé tartoznak a következők: jitter, shimmer és Harmonics to Noise Ratio (HNR). Korábbi kutatások igazolják, hogy olyan akusztikai paraméterek, mint a jitter, shimmer, HNR és az MFC-együtthatók (Mel-frequency cepstral coefficients) első komponense (c_1) (a továbbiakban: mfcc01) az egészséges és kóros hangok automatikus osztályozásában hasznos paraméterek [4,5,6].

Korábban kimutattuk [7], hogy ezek az akusztikai paraméterek magas korrelációt mutatnak a diszfónia súlyosságával, ugyanúgy, mint a különböző fonetikai osztályokon mért Soft Phonation Index (SPI) és Empirical mode decomposition (EMD) alapú frekvenciasáv-arányok. Ezek a paraméterek hasznosak lehetnek a diszfónia különböző típusainak, például funkcionális diszfónia és recurrens paresis megkülönböztetéséhez.

A diszfónikus beszéd diagnózisát és kezelését a páciens hangminőségét vizsgáló orvos végzi. A beszéd értékelése természeténél fogva szubjektív. A diszfóniás hang súlyosságát általában egy orvos értékelése, vagy több tapasztalt értékelő ítéletének átlaga vagy mediánja alapján határozzák meg [8,9]. Ha több értékelő áll rendelkezésre, a szakemberek a diszfónia súlyosságának értékelését korábban rögzített hangminták meghallgatásával végzik. Az értékelések változhatnak az értékelő szakemberek között, ezért tanácsos konzisztenciaelemzést végezni.

Ebben a tanulmányban organikus és funkcionális diszfónia súlyosságának automatikus becslésére összpontosítunk. Egy diszfónia súlyosságát becslő módszert javaslunk, amely magyar anyanyelvű betegek és kontrollpopuláció felolvasott szövegein alapul. A diszfónia súlyosságát négy szakember határozta meg: a szakember, aki a pácienst kezelte és közvetlenül értékelte a páciens beszédének minőségét a konzultáció során, valamint további három szakember, aki nem ismerte a pácienseket, csak az előzetesen rögzített hangfájlokat hallgatta meg és azok alapján határozta meg a diszfónia súlyosságát. A diszfónia súlyosságát az RBH-skála szerint határozták meg a szakemberek, ahol az R az érdességet, a B a levegősséget, a H pedig a rekedtséget jelenti [10]. Korábbi eredményeink alapján [7] 18 különböző fonetikai osztályon (például nazálisokon, magas magánhangzókon, zöngés spiránsokon stb.) mért akusztikai paramétert választottunk ki és végeztünk osztályozási, valamint regressziós elemzéseket. Négyosztályú osztályozáshoz k-közép felügyelet nélküli osztályozót használtunk, annak érdekében, hogy megtudjuk a használt akusztikai paraméterek alkalmasak-e a szakemberek négyfokozatú értékelésének modellezésére. Megvizsgáltuk az RBH szubjektív jellegét, valamint a négy szakember minősítésének konzisztenciáját. A pácienseket kezelő szakem-

ber értékelését összehasonlítottuk azon szakemberek értékeléseivel, akik az előre rögzített hangfelvételek alapján határozták meg a diszfónia súlyosságát.

A Bevezetés után, a 2. fejezetben röviden ismertetjük a kísérletekben használt beszédadatbázist, a mért akusztikai paramétereket, valamint a kiértékelési módszereket. Eredményeinket a 3. fejezetben mutatjuk be, majd a 4. fejezetben ismertetjük az eredmények interpretálását és a jövőbeli terveinket.

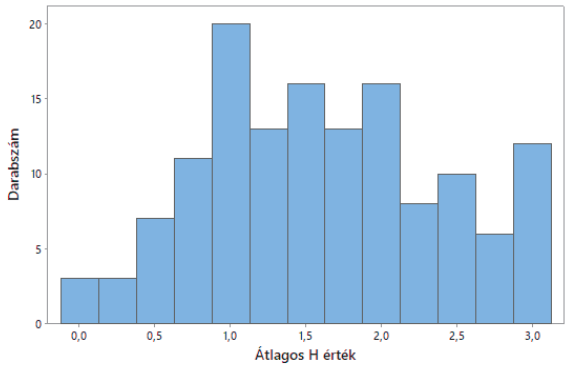
2. Módszerek és beszédanyagok

2.1. Patológiás és egészséges beszédadatbázis

A páciensek hangmintáinak felvétele, az Országos Onkológiai Intézet Fül-orr-gégészeti osztályának járóbeteg ellátásai során történt. A vizsgálat során számos betegség fordult elő: funkcionális diszfónia, recurrens paresis (hangszalagbénulás), a hangképző szervrendszer különböző pontjain előforduló tumorok, gasztro-esophageal reflux (GERD), krónikus gégegyulladás, bulbar paresis (agyideggyulladás), amiotrófiás laterálszklerózis (ALS), leukoplakia, spazmodikus diszfónia, stb. Összehasonlítás végett egészséges páciensekről is készültek felvételek. Ezek rögzítése a korábban említett esetekhez nem kapcsolódó kivizsgálások során kerültek felvételre.

2.2. Az RBH skála

A hang kórosságának súlyosságát négy szakember határozta meg: a szakorvos, aki kezelte a pácienst, felállította a diagnózist és meghatározta a hangjának súlyosságát a konzultációk során (Szakember 1), valamint további három szakember (Szakember 2, Szakember 3 és Szakember 4) aki nem került személyes kapcsolatba a páciensekkel, csupán a hangfelvételek visszahallgatása alapján határozták meg a diszfónia súlyosságát. Mind a négy szakembernek nagy tapasztalata van a hangképzési rendellenességekkel rendelkező páciensek kezelésében. A diszfónia súlyosságát az RBH skálával lehet meghatározni, ahol az R (roughness) az érdességet, a B (breathiness) a levegősséget, míg a H (hoarseness) az általános rekedtséget hivatott mutatni. A H értéke nem lehet kisebb, mint a másik két kategória maximuma. Például ha a B=3 és az R=2, akkor a H értéke 3, nem lehet 2, vagy 1. Az egészséges hangok kódja így R0B0H0 szerint alakul. Ptok és társai igazolták, hogy az RBH skála alkalmazása megfelelő a klinikai alkalmazásokra [11]. A tanulmányunk során az általános rekedtség (H) vizsgálata, és diagnosztizálása történt. 138 felvétel került felhasználásra. Minden felvétel esetén a szakemberek külön külön meghatározták a diszfónia súlyosságának fokát az RBH skála szerint, majd a H értékeket felvételenként átlagoltuk, ezt használtuk fel később célváltozóként. A szakemberek által meghatározott H értékek átlagának eloszlása az adatbázisban az 1. ábrán látható.



1. ábra: A H érték eloszlása az adatbázisban.

2.3. A felvételek rögzítési környezete és szövege

A felvételek (Monacor ECM-100) közeltéri mikrofonnal, alacsony zajszinttű külső hangkártyával (Creative Soundblaster Audigy 2 NX), jó minőségű A/D konverterrel (kódolás: PCM, mintavételezési frekvencia: 16 kHz, kvantálás: 16-bit) kerültek rögzítésre, csendes irodai környezetben (orvosi szobában). Minden páciens Aiszóposz meséjét, „Az északi szél és a nap" -ot olvasta fel. Ezen népmese gyakran használt a foniátriai kutatásokban, mivel a mese fonetikailag kiegyensúlyozott. Vagyis a szöveganyagát úgy szerkesztették meg, hogy az adott nyelvben előforduló minden beszédhang, valamint a leggyakoribb hangkapcsolatok szerepelnek benne. Számos nyelvre elkészült ez a szöveg, köztük a jelen esetben is használt magyarra. Az adatbázis fonéma szintű szegmentálása automatikus fonémaszegmentáló programmal történt, amit a Beszédakusztikai Laboratórium munkatársai fejlesztettek ki [12], majd szükség esetén kézileg történt ezek javítása.

2.4. Akusztikai paraméterek

A korábbi munkák során az alap paraméter szett a következő volt: jitter(ddp), shimmer(ddp), Harmonics-to-Noise Ratio (HNR) és mfcc01, melyeknek átlaga (mean), illetve szórása (std) a népmesében legtöbbet szereplő [E] (SAMPA) magánhangzóból kerültek kinyerésre [5] (továbbiakban Baseline paraméter szett). Ezt követően a paraméterek listája további két akusztikai mérőszámmal bővült, a Soft Phonation Index (SPI)-el és az Empirical mode decomposition (EMD) módszeren alapuló frekvencia sávok arányaival, melyeket különböző fonetikai osztályokba tartozó beszédhangokon vizsgáltuk meg.

Az SPI a beszéd jel alacsony frekvencia sávjának (70 - 1600 Hz) és a magas frekvencia (1600 - 4500 Hz) sávjának energiaaránya. Ha ez az arány magas, akkor az azt jelenti, hogy az energia az alacsony frekvencia sávra koncentrálódik, amely lágyabb hangot eredményez [13].

Fogalmilag az EMD (Empirical Mode Decomposition) a multikomponenses jelet szétbontja elemi jel komponensekké, amik aztán úgy ismertek, mint Intrinsic mód függvények (Intrinsic Mode Functions - IMF). Az IMF-ek frekvencia szerint vannak mátrixba rendezve. Az első pár IMF a magas frekvenciájú komponense a jelnek, míg a továbbiak a kisebb frekvenciájú komponenseket reprezentálják. Minden IMF esetén kiszámításra került az entrópiája (H). A frekvencia sávok entrópiáinak aránya pedig a következő módon került meghatározásra:

$$IMF_{entropy} = \frac{\sum_{d=1}^{2} H_d}{\sum_{d=2}^{D} H_d} \tag{1}$$

H_d a log transzformált IMF-ek Shannon entrópia értéke az egyes $d = 1, 2, \ldots D$ értékekre, ahol D az összes kinyert IMF száma. A Shannon entrópia diszkrét jelek esetén a következőképpen van definiálva:

$$H(p_i) = -K \sum_{i=1}^{n} p_i log p_i \tag{2}$$

Ahol K pozitív konstans. Az IMF paraméter meghatározása a [14] Matlab toolkittel történt.

Az SPI, valamint az IMF entrópia értékét nem csak az [E] magánhangzóra, hanem további fonetikai csoportokra is meghatároztuk. A vizsgálat során az SPI és az IMF entrópiáját a következő hat fonetikai osztályra számoltuk ki:

- az [E] magánhangzón
- nazális hangokon (Nasal), mint [m], [n] és [J]
- magas magánhangzókon (HighVowels), mint [E], [e:], [i], [2] és [y]
- mély magánhangzókon (LowVowels), mint [O], [A:], [o] és [u]
- spiránsokon (VoicedSpirants), mint [v], [z] és [Z]
- zárhangokon és affrikátákon (VoicedPlosives), mint [b], [d], [g], [dz], [dZ] és [d']

Az SPI paramétert az egész hangfájlra is kinyertük, a többi esetben felvételenként átlag és szórás is kiszámításra került. A SPI és az IMF akusztikai paramétereken kívül, a jitter, shimmer, HNR és az mfcc01 felvételenkénti átlag- és szórásértékeivel összesen 33 paraméter került vizsgálatra.

2.5. Döntési módszerek

Felügyelet nélküli klaszterelemzéssel megvizsgáltuk, hogy a különböző fonetikai osztályokon mért akusztikai paraméterek alkalmasak-e a szakemberek négyfokozatú (RBH szubjektív skála szerinti) értékelésének modellezésére. Ehhez k-közép klaszterező algoritmust használtunk. A k-közép az egyik legegyszerűbb algoritmus, amely felügyelet nélküli tanulási módszert alkalmaz az ismert klaszterezési

problémák megoldására. Ez az eljárás egyszerű és gyors megközelítése a problémának, könnyen implementálható és könnyen értelmezhető az eredménye.

Megvizsgáltuk a négy szakember értékeléseinek konzisztenciáját. A diszfónia súlyosságának becslésére lineáris regressziót és radiális bázisfüggvényű (RBF) SVR-t (Support Vector Regression, továbbiakban SVR) használtunk. A lineáris regresszió természetéből adódóan csak az egymástól függő és a független változók lineáris kapcsolatát vizsgálja; továbbá, azt is feltételezi, hogy a bemeneti változók és a célváltozó között lineáris kapcsolat van. Az RBF-kernellel rendelkező SVR-nek jó általánosító képessége van és robosztus a bemenő zajra tekintettel.

Az egyenletesebb eloszlás érdekében az adatbázist kibővítettük 10 extra egészséges felvétellel (H0) a klaszter- és a regresszióelemzés feladatokhoz. E populáció beszédét ugyanazzal a szöveges anyag felhasználásával rögzítettük. Az páciensek nem rendelkeztek más betegséggel, és nem voltak semmilyen orvosi kezelés alatt. A következetesség vizsgálat során az eredeti 138 mintát használtuk.

3. Eredmények

3.1. Felügyelet nélküli klaszterelemzés

Négy osztályos k-közép klaszter analízist végeztünk egy 18 paramétert tartalmazó bemenő jellemzővektorral, amely jó eredményekkel tudta osztályozni az egészséges és diszfóniás hangokat. A 18 paraméteres bemenő jellemzővektor egy Forward Feature Selection (FFS) algoritmus eredményeként került kiválasztásra. Az FFS egy iteratív algoritmus, amely kiválasztja a legjobb paramétert egy előre meghatározott költségfüggvény kielégítése alapján, minden lépésben egy újabb paramétert hozzáadva a paraméterek halmazába. Jelen esetben a paraméterek a legalacsonyabb RMSE érték szerint lettek kiválasztva. A kiválasztott paraméterek az 1. táblázat tartalmazza.

1. táblázat. 18 akusztikai paraméteres szett.

	Kiválasztott akusztikai paraméterek
18 paraméter szett	jitter$_{mean}$, shimmer$_{mean}$, hnr$_{mean}$, mfcc01$_{mean}$, jitter$_{std}$, shimmer$_{std}$, hnr$_{std}$, mfcc01$_{std}$, SPI->E$_{std}$, SPI->Nasal$_{mean}$, SPI->Nasal$_{std}$, SPI->LowVowels$_{std}$, SPI->VoicedSpirants$_{mean}$, SPI->VoicedSpirants$_{std}$, IMF->E$_{std}$, IMF->Nasal$_{mean}$, IMF->VoicedPlosives$_{mean}$, IMF->VoicedPlosives$_{std}$

Érdekes kérdés, hogy ezen kiválasztott paraméterek modellezni tudják-e az egyéni értékeléseket. Ha az akusztikai paraméterek szettje, és a felügyelet nélküli tanuló algoritmus rögzített, akkor összehasonlítható a négy szakember értékeléseinek klaszter modellje külön-külön. Az RBH skála szubjektív voltának vizsgálata céljából klaszter analízist végeztünk.

Az egyes szakemberek tévesztési mátrixait a 2., 3., 4. és 5. táblázat reprezentálja. Az osztályozások teljesítményét a 6. táblázat mutatja. A H döntések

2. táblázat. Tévesztési mátrix Szakember 1 esetén.

		Prediktált H				
		0	1	2	3	Összesen
Szakember 1 által megállapított H érték	0	**12**	13	9	2	36
	1	1	**33**	25	4	63
	2	2	5	**10**	6	23
	3	1	3	5	**17**	26
	Összesen	16	54	49	29	148

3. táblázat. Tévesztési mátrix Szakember 2 esetén.

		Prediktált H				
		0	1	2	3	Összesen
Szakember 2 által megállapított H érték	0	**11**	5	6	0	22
	1	3	**26**	24	2	55
	2	2	23	**16**	15	56
	3	0	0	3	**12**	15
	Összesen	16	54	49	29	148

4. táblázat. Tévesztési mátrix Szakember 3 esetén.

		Prediktált H				
		0	1	2	3	Összesen
Szakember 3 által megállapított H érték	0	**11**	2	3	0	16
	1	2	**20**	15	1	38
	2	2	25	**16**	9	52
	3	1	7	15	**19**	42
	Összesen	16	54	49	29	148

5. táblázat. Tévesztési mátrix Szakember 4 esetén.

		Prediktált H				
		0	1	2	3	Összesen
Szakember 4 által megállapított H érték	0	**12**	7	6	0	25
	1	2	**24**	18	6	50
	2	2	18	**17**	6	43
	3	0	5	8	**17**	30
	Összesen	16	54	49	29	148

6. táblázat. Teljesítménymetrikák minden szakember és címke esetén.

	Metrika	**H0**	**H1**	**H2**	**H3**
Szakember 1 által megállapított H érték	Precision	0,33	0,52	0,43	0,65
	Recall	0,75	0,61	0,20	0,59
	F-score	0,46	0,56	0,28	0,62
Szakember 2 által megállapított H érték	Precision	0,50	0,44	0,41	0,80
	Recall	0,69	0,44	0,47	0,41
	F-score	0,58	0,44	0,44	0,55
Szakember 3 által megállapított H érték	Precision	0,69	0,53	0,31	0,45
	F-score	0,69	0,43	0,32	0,54
Szakember 4 által megállapított H érték	Precision	0,48	0,48	0,40	0,57
	Recall	0,75	0,44	0,35	0,59
	F-score	0,59	0,46	0,37	0,58

átlagos pontossága az egyes szakemberek esetén sorrendben: 0,49; 0,44; 0,45; 0,47.

Ebből a vizsgálatból azt a következtetést vonhatjuk le, hogy a bemenő jellemzővektor alkalmas a diszfónia súlyosságának egyedi értékelésére. Összességében elmondható az egyes tévesztési mátrixok alapján, hogy amennyiben a klaszterezési eljárás nem tudta pontosan meghatározni a szakértő értékelését, akkor is főként a szomszédos klaszterekbe sikerült besorolnia az adott felvételt. Az egészséges (H0) és a patológiás (H1, H2 és H3) hangok elválasztása kielégítő. A tévesztési mátrixok megmutatják az egyes szakemberek minősítési stílusát, miszerint míg Szakember 1 a hangokat kevésbé ítélte súlyosnak, addig Szakember 4 több H3-as súlyosságú értékelést adott. Annak ellenére, hogy a Szakember 3 értékelései voltak a legkevésbé pontosak, mégis a H3 értékeléseinek pontossága meglepően magas, 0,80 volt. Az eredmények azt is mutatják, hogy mind a négy szakember értékelései esetén a H1 és H2-es esetek állnak legközelebb egymáshoz.

Ez jól mutatja, hogy jelen esetben egy folytonos skálát becslő eljárás, mint a regresszió, jobban tudná közelíteni az értékeléseket, mint a diszjunkt halmazokkal dolgozó klaszterezés, így pontosabb, kisebb hibával dolgozó rendszert eredményezve.

3.2. Következetesség vizsgálat

Ebben a részben az értékelések következetességének vizsgálata kerül bemutatásra, valamint, hogy megfigyelhető-e különbség a felvételek készítésénél ott lévő szakember és a hangfájlokat visszahallgató további három szakember értékelései között. A 7. táblázaton látható, hogy a szakemberek átlagos eltérése a H átlagától 0,4. Eszerint a Szakember 2 és 4 ad leginkább átlagos értékeléseket, így az ő értékelései változtatják meg legkevésbé az átlagos H értéket, míg Szakember 1 értékelései különböznek leginkább az átlagtól. Külön kiemelendő viszont, hogy egyedül Szakember 1 volt jelen a felvételek elkészítésénél.

A Cronbach alfa a leggyakrabban használt mérőszám a belső konzisztencia mérésére, kifejezésére. A 0,8 vagy afölötti Cronbach alfa értékek magas szintű belső konzisztenciát indikálnak. Jelen esetben a Chronbach alfa 0,891-nek adódott, ami jó belső konzisztenciát mutat. Ezen Chronbach értéket egyik szakember kihagyása sem növelné. Megjegyzendő, hogy Szakember 1 értékelésének törlésével csökkenne legkisebbet a belső konzisztencia, vagyis a többi szakember törlése esetén az jelentősebben csökkenne. Az eredményeket a 8. táblázat foglalja össze.

Az Intra Class Correlation Coefficient (ICC) is megállapításra került. Az ICC 0,75 feletti értéke kiváló, míg az alatti értéke jó következetességet mutat. Az értékelésekre jó konzisztencia adódott, míg a szakemberek értékeléseinek együttes (átlagos) konzisztenciája 0,89 lett, amely kiváló konzisztenciát mutat, 95% konfidencia intervallum esetén 0,857-tól 0,917-ig ($F(137;411) = 9{,}172$; $p < 0{,}001$).

3.3. Regresszióelemzés

A regresszió jelentős előnyt jelent a klaszterelemzéshez képest, mivel a becslés szinte folyamatosan követi a függvényt. Ez a tulajdonság jelentősen javíthatja

7. táblázat. A H átlagtól való átlagos eltérések az egyes szakembereknél.

	Összesen	Szakember 1	Szakember 2	Szakember 3	Szakember 4
H átlagtól való átlagos eltérések	0,4	0,5	0,3	0,4	0,3

8. táblázat. Megbízhatósági statisztika, ha valamelyik tag törölve van.

	Elem, ha törölt	
	item-rest korreláció	**Cronbach's α**
Szakember 1	0,714	0,881
Szakember 2	0,777	0,857
Szakember 3	0,782	0,852
Szakember 4	0,787	0,850

a modell minőségét. A regressziós módszerek jóságát, a négyzetes középérték hiba (RMSE), az R^2 (a determináció együtthatója, azaz a függő változó varianciájának aránya, amelyet a független változók magyarázhatnak) és a Pearson korreláció (a cél és a prediktált H értékek között) értékével vizsgáltuk. Ebben az elemzésben lineáris regressziót és RBF kernelfüggvényű SVR regressziót használtunk.

A legjobb modell elérése érdekében több paraméter szett is kipróbálásra került bemenő jellemzővektorként. Összehasonlításra került a kiindulási paraméter szett, a klaszteranalízisben használt 18 paraméteres szett, valamint az FFS algoritmus kimenete lineáris és RBF kernelfüggvényű regresszió esetén külön. Az RBF kernelt használó regressziónál a hiperparaméter-optimalizáció grid keresésel történt.

9. táblázat. Az FFS algoritmus eredménye, a négy szakember értékelésének átlaga célváltozóként.

	Kiválasztott akusztikai paraméterek
FFS lineáris kernellel	mfcc01_{mean}, shimmer_{mean}, SPI->LowVowels_{std}, hnr_{std}, SPI->HighVowels_{mean}, IMF->Nasal_{mean}, SPI->$\text{VoicedPlosives}_{std}$, IMF->$\text{DeepVowels}_{std}$
FFS RBF kernellel	shimmer_{mean}, hnr_{mean}, mfcc01_{mean}, hnr_{std}, SPI->E_{mean}, SPI->E_{std}, SPI->Nasal_{std}, SPI->HighVowels_{mean}, SPI->LowVowels_{mean}, SPI->LowVowels_{std}, SPI->$\text{VoicedPlosives}_{mean}$, IMF->$\text{Nasal}_{mean}$, IMF->$\text{VoicedPlosives}_{mean}$, IMF->$\text{VoicedPlosives}_{std}$

A 11. táblázat mutatja a regresszióelemzés eredményét, amikor a négy szakember értékelésének átlaga volt a célváltozó. Az FFS algoritmus eredményeit, lineáris és RBF-kernel esetekben a 9. táblázat foglalja össze. A lineáris regressziót használó FFS adta a legnagyobb, 0,853-as korrelációt, viszont az RBF kernel használatával érhető el a H legkisebb RMSE értéke (0,454). A legkisebb korrelációt és a legnagyobb RMSE hibát adó jellemzővektor a Baseline paramétereket használó összeállítás volt. Ezek alapján megállapítható, hogy a különböző fonetikai osztályokba tartozó beszédhangokon mért akusztikai paraméterek (SPI és

IMF entropy energiahányadosok) növelik a hangképzési rendellenességek súlyosságát prediktáló modellek jóságát. Az FFS kimenetével képzett jellemzővektorok modelljei csak kis mértékben térnek el a 18 paraméteres jellemzővektoros modelltől, így ez is mutatja annak általános jóságát a probléma esetén.

A Szakember 1 predikcióira is modell épült, melynek eredményeit a 12. táblázat mutatja, mivel ő volt jelen a felvételek elkészítésénél, ő hallotta élőben a pácienst, és állította fel a diagnózist. Az FFS algoritmus eredményeit, lineáris és RBF-kernel esetekben a 10. táblázat foglalja össze.

A 12. táblázat azt mutatja, hogy a legjobb eredményt a lineáris regresszióval kapott paraméterkiválasztás adta, a legnagyobb 0,759-es korrelációval, és a legkisebb 0,687-es RMSE hiba értékkel. Annak ellenére, hogy a korreláció 0,1-el kisebb, és a hiba értéke 0,2-vel nagyobb az előzőhöz képest, fontos különbség van a két eset között: a súlyossági osztály felosztása. Amíg az előző esetben a négy szakember átlaga 0,25-ös felbontású skálát ad, addig itt az egyetlen szakember értékelése esetén 1-es a skála felbontása a rekedtség súlyosságára. Ennek ellenére azonos következtetés vonható le jelen eredmények kapcsán is, miszerint a különböző fonetikai osztályokon mért energiahányadosok (SPI and $\text{IMF}_{entropy}$) növelni tudják a felállítható modell jóságát.

10. táblázat. Az FFS algoritmus eredménye, Szakember 1 értékelése célváltozóként.

	Kiválasztott akusztikai paraméterek
FFS lineáris kernellel	hnr_{mean}, IMF->Nasal_{mean}, shimmer_{mean}, mfcc01_{mean}, SPI->E_{mean}, SPI->E_{std}, hnr_{std}, SPI->$\text{VoicedPlosives}_{std}$, SPI->$\text{LowVowels}_{std}$, IMF->$\text{VoicedSpirants}_{std}$
FFS RBF kernellel	shimmer_{mean}, hnr_{mean}, mfcc01_{mean}, shimmer_{std}, hnr_{std} SPI->file, SPI->E_{mean}, SPI->Nasal_{mean}, SPI->Nasal_{std}, SPI->HighVowels_{mean}, SPI->LowVowels_{mean}, SPI->LowVowels_{std}, SPI->$\text{VoicedSpirants}_{mean}$, SPI->$\text{VoicedPlosives}_{mean}$, SPI->$\text{VoicedPlosives}_{std}$, IMF->$\text{E}_{mean}$, IMF->$\text{Nasal}_{mean}$, IMF->$\text{Nasal}_{std}$, IMF->$\text{HighVowels}_{std}$, IMF->$\text{LowVowels}_{mean}$, IMF->$\text{LowVowels}_{std}$, IMF->$\text{VoicedSpirants}_{std}$, IMF->$\text{VoicedPlosives}_{mean}$

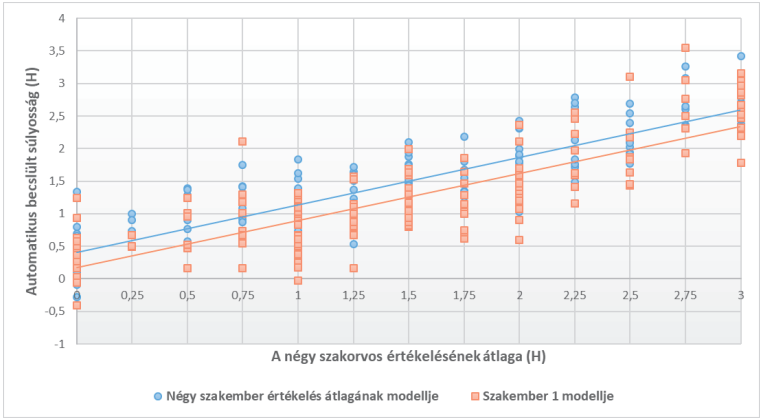
11. táblázat. Regresszió elemzés eredménye, a négy szakember értékelés átlaga célváltozóként.

Akusztikai paraméter szett	**Regresszió típusa**	**R^2**	**Korreláció**	**H RMSE értéke**	**Hiperparaméterek**
Baseline paraméter szett	Lineáris	0,639	0,799	0,531	-
18 paraméter szett	Lineáris	0,700	0,837	0,502	-
FSS eredménye	Lineáris	**0,727**	**0,853**	0,462	-
Baseline paraméter szett	RBF kernel	0,660	0,812	0,500	$C = 8$, $\gamma = 0{,}25$
18 paraméter szett	RBF kernel	0,653	0,808	0,506	$C = 2$, $\gamma = 0{,}125$
FSS eredménye	RBF kernel	0,720	0,849	**0,454**	$C = 4$, $\gamma = 0{,}25$

12. táblázat. Regresszió elemzés eredménye, Szakember 1 értékelése célváltozóként.

Akusztikai paraméter szett	Regresszió típusa	R^2	Korreláció	H RMSE értéke	Hiperparaméterek
Baseline paraméter szett	Lineáris	0,476	0,690	0,759	-
18 paraméter szett	Lineáris	0,544	0,738	0,735	-
FSS eredménye	Lineáris	**0,577**	**0,759**	**0,687**	-
Baseline paraméter szett	RBF kernel	0,413	0,643	0,786	$C = 4$, $\gamma = 0{,}0625$
18 paraméter szett	RBF kernel	0,420	0,648	0,775	$C = 2$, $\gamma = 0{,}125$
FSS eredménye	RBF kernel	0,504	0,710	0,716	$C = 4$, $\gamma = 0{,}25$

A 2. ábra mutatja a diszfónia automatikusan prediktált súlyosságát az eredeti perceptuális referencia értékekhez képest, mind a két regressziós modell esetén: a négy szakember értékelésének átlaga alapján készített és a betegeket kezelő klinikus ítélete alapján létrehozott modell. Az ábra az FFS algoritmussal kapott lineáris regresszió modellt mutatja, mindkét esetben.



2. ábra: Automatikusan becsült diszfónia súlyosság a perceptuálisan megítélt H érték alapján.

A 2. ábra ismételten bemutatja a javasolt megközelítés becslési hatékonyságát, függetlenül a páciens patológiás hátterétől és a diszfóniájának súlyosságától. A két modell hasonló predikciókat ad, viszont Szakember 1 modellje nagyobb varianciával becsül, valamint ezen modell kisebb súlyossági fokúnak prediktálta a hangokat, mint a négyértékeléses modell. Látható, hogy mind a két modell jó predikciót ad a H1 súlyossági osztályra, súlyosabb esetekben pedig alulról becsli azokat.

4. Következtetés

Jelenlegi tanulmány a diszfónia súlyosságának automatikus becslésére irányult, a szakemberi értékelések szubjektív jellegének figyelembevételével. Ez a kutatás fo-

lyamatos beszéden alapult, mivel az jobban alkalmazható a gyakorlati munkára, mint a kitartott magánhangzók [2].

Az összes beszédfelvételt négy szakértő értékelte: a szakorvos, aki kezelte a pácienst, felállította a diagnózist és meghatározta a hangjának súlyosságát a konzultációk során, valamint további három szakember, aki nem került személyes kapcsolatba a páciensekkel, csupán a hangfelvételek visszahallgatása alapján határozták meg a diszfónia súlyosságát. A diszfónia súlyosságát az RBH skála szerint határozták meg, ebben a tanulmányban az általános rekedtséget (H) vizsgáltuk.

Az eredményeink azt mutatják, hogy a különböző fonetikai osztályokon mért energiahányadosok (SPI és $\mathrm{IMF}_{entropy}$) növelni tudják a felállítható modell jóságát, vagyis hasznosak a hangképzési rendellenességek súlyosságának predikciójában.

A k-közép, felügyelet nélküli tanulási módszer segítségével négy modellt hasonlítottunk össze, külön a négy szakember értékelése szerint. A kiválasztott akusztikai paraméterek alkalmasak a szakemberek négyfokozatú értékelésének modellezésére. A H döntés pontossága a négy szakember szerint 0,49; 0,44; 0,45; 0,47. A diszfónia súlyossági ítéletek között nagyfokú megbízhatóság volt tapasztalható a Cronbach Alpha és ICC-vel való belső konzisztenciájának mérésekor.

A diszfónia súlyosságának automatikus becslése során két regressziós modellt hasonlítottunk össze: a négy szakember értékelésének átlaga által készített modellt és a betegeket kezelő klinikus külön modelljét. A két modell alacsony RMSE és magas korrelációs értékeket ért el az automatikusan becsült súlyosság és perceptuális értékelések között.

A jövőbeni munka magában foglalja további szakemberek bevonását a diszfónia súlyosságának megítélésében, valamint az adatbázis bővítését. Egy nagyobb adatbázis lehetővé teheti különböző diszfónia típusok osztályozását. Úgy gondoljuk, hogy a tanulmányban bemutatott módszer általánosítható más nyelvekre is.

Hivatkozások

1. Hirschberg, J., Hacki, T., Mészáros, K.: Foniátria és társtudományok: A hangképzés, a beszéd és a nyelv, a hallás és a nyelés élettana, kórtana, diagnosztikája és terápiája (I. kötet). Budapest: Eötvös Kiadó. (2013)
2. Kim, J., Kumar, N., Tsiartas, A., Li, M., Narayanan, S.S.: Automatic intelligibility classification of sentence-level pathological speech. Computer speech & language **29**(1) (2015) 132–144
3. Liu, Y., Lee, T., Ching, P., Law, T.K., Lee, K.Y.: Acoustic assessment of disordered voice with continuous speech based on utterance-level asr posterior features. Proc. Interspeech 2017 (2017) 2680–2684
4. Vicsi, K., Imre, V., Mészáros, K.: Voice disorder detection on the basis of continuous speech. In: 5th European Conference of the International Federation for Medical and Biological Engineering, Springer (2011) 86–89
5. Kazinczi, F., Mészáros, K., Vicsi, K.: Automatic detection of voice disorders. In: International Conference on Statistical Language and Speech Processing, Springer (2015) 143–152

6. Grygiel, J., Strumiłło, P., Niebudek-Bogusz, E.: Application of mel cepstral representation of voice recordings for diagnosing vocal disorders. Delta **12** (2012) 2
7. Tulics, M.G., Vicsi, K.: Phonetic-class based correlation analysis for severity of dysphonia. In: Cognitive Infocommunications (CogInfoCom), 2017 8th IEEE Conference on, IEEE (2017) 21–26
8. Chien, Y.R., Borskỳ, M., Guðnason, J.: Objective severity assessment from disordered voice using estimated glottal airflow. Proc. Interspeech 2017 (2017) 304–308
9. Laaridh, I., Kheder, W.B., Fredouille, C., Meunier, C.: Automatic prediction of speech evaluation metrics for dysarthric speech. Proc. Interspeech 2017 (2017) 1834–1838
10. Schönweiler, R., Hess, M., Wübbelt, P., Ptok, M.: Novel approach to acoustical voice analysis using artificial neural networks. JARO-Journal of the Association for Research in Otolaryngology **1**(4) (2000) 270–282
11. Ptok, M., Schwemmle, C., Iven, C., Jessen, M., Nawka, T.: On the auditory evaluation of voice quality. HNO **54**(10) (2006) 793–802
12. Kiss, G., Sztaho, D., Vicsi, K.: Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In: Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on, IEEE (2013) 579–582
13. Roussel, N.C., Lobdell, M.: The clinical utility of the soft phonation index. Clinical linguistics & phonetics **20**(2-3) (2006) 181–186
14. Tsanas, A.: Acoustic analysis toolkit for biomedical speech signal processing: concepts and algorithms. Models and analysis of vocal emissions for biomedical applications **2** (2013) 37–40